Abstract

The school-based Playing-2-gether is a 12-week intervention with two components aimed at decreasing child externalizing behavior through improving teacher-child interactions. The first component is rooted in attachment theory and aimed at enhancing teacher-child relationship quality, and the second is based on learning theory and aimed at improving teachers' behavior management. In this three-wave randomized study, effects of Playing-2-gether on the teacher-child relationship quality and on teacher-rated child behavioral adjustment were investigated. To this aim, 175 dyads consisting of male preschoolers with relatively high levels of externalizing problem behavior and their teachers were randomly assigned to Playing-2-gether (n = 89) or an education-as-usual control condition (n = 86). Teacher-rated questionnaires were collected at pre-test, after the first intervention component, and at post-test. At post-test, the intervention group showed a larger decrease in teacher-child conflict, child conduct problems, and child hyperactivity/inattention. Supplementary analyses showed that all positive effects were already visible after the first intervention component and that teacher-child conflict, child conduct problems and hyperactivity/inattention did not further reduce during the second component. In addition, an increase in closeness was found following the first component, but subsequently disappeared at post-test.

Key words: intervention, teacher-child relationship, externalizing problem behavior, problem behavior, behavioral adjustment, preschool

Improving Teacher-Child Relationship Quality and Teacher-Rated Behavioral Adjustment Among Externalizing Preschoolers: Effects of a Two-Component Intervention

Externalizing problem behavior (EPB) in early childhood refers to a range of behaviors that are disruptive and/or harmful for others, such as overactive, oppositional, and aggressive behavior (Smidts & Oosterlaan, 2007). This pattern of behavior is predictive for maladjustment later in life, such as conduct disorder, school failure, and mental disorders. However, not all young children showing EPB develop problems (e.g., Dodge, Coie, & Lynam, 2006). To prevent EPB later in life, it is important that we learn more about the factors associated with the stability and malleability of preschoolers' EPB.

In addition to parent-child interactions, teacher-child interactions are increasingly considered as environmental factors that influence preschooler behavioral adjustment (Pianta, Hamre, & Stuhlman, 2003). More specifically, a growing number of observational studies have focused on the link between the affective quality of the teacher-child relationship (TCR) and teacher behavior management, on the one hand, and preschooler behavioral adjustment, on the other (e.g., Cowan & Sheridan, 2009; Pianta et al., 2003). Also, interventions grounded in attachment and learning theory, aimed at improving preschooler behavioral adjustment by focusing on both enhancing the TCR and teacher behavior management, are on the rise (e.g., Gershenson, Lyon, & Budd, 2010; Lyon et al., 2009; Webster-Stratton & Reid, 2009). The main goal of the present study is to contribute to this body of literature by investigating the effect of a two-component intervention (i.e., Playing-2-gether) on the TCR and preschooler behavioral adjustment in a randomized controlled trial. Moreover, this study aims to improve insight in what happens during this intervention by including an intermediate assessment after the first, relationship-focused component.

Attachment and Learning Theory and Teacher-Child Adjustment

According to *attachment theory*, developing a secure attachment bond with primary caregivers is important to promote (young) children's development and to prevent adjustment problems in general and EPB in particular (see meta-analysis by Fearon, Bakermans-Kranenburg, van Ijzendoorn, Lapsey, & Roisman, 2010). As teachers can be considered as temporary or ad hoc attachment figures playing the role of secure base and safe haven (e.g., Verschueren & Koomen, 2012), they may also impact preschoolers' adjustment.

More specifically, from an attachment perspective, the TCR has been conceptualized along positive and negative affective dimensions, which have been labeled closeness (i.e., the amount of warm interactions and open communication) and conflict (i.e., the discordance and negativity in the relationship; Pianta et al., 2003). Closeness refers to proximity seeking behavior of the child and the degree to which the teacher is used as a safe haven versus avoided in stressful situations. Conflict refers to resistance and disharmony in insecure TCR's (Verschueren & Koomen, 2012). It has been shown that teacher-child closeness can have a moderating, compensatory effect on EPB development in preschoolers with initially elevated EPB levels (e.g., Silver, Measelle, Armstrong, & Essex, 2005). On the other hand, several longitudinal studies revealed that teacher-child conflict in preschool is linked with increased EPB levels (e.g., Doumen et al., 2008).

In sum, high-quality teacher-child relationships are considered important resources that may help children to maintain adequate behavior or redirect EPB in times of stress at school (e.g., Driscoll & Pianta, 2010). In building high-quality teacher-child relationships, teacher sensitivity plays a key role (e.g., Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008). To convey sensitivity to the child's needs, it is important that teachers observe, describe, and label feelings of children correctly (Driscoll & Pianta, 2010). Based on these premises, Banking Time (Pianta & Hamre, 2001) was developed, which is a dyadic teacher-child intervention consisting of child-centered play sessions. Banking Time has been shown to be

effective in improving teacher-child closeness (not conflict) and teacher-reported behavioral adjustment (Driscoll & Pianta, 2010; Driscoll, Wang, Mashburn, & Pianta, 2011).

Learning theory focuses on the importance of teacher-child interactions for children's behavioral development as well. More specifically, the operant conditioning model stresses the importance of the antecedents and the consequences of behavior as targets for behavioral change (cf. ABC-model, see Cowan & Sheridan, 2009). By setting up the conditions under which the desired behavior is likely to occur (i.e., antecedent manipulation) and by reinforcing the behavior if it occurs (i.e., consequence manipulation), the incidence of desired behavior can be enhanced. Alternatively, punishment can be used to reduce undesired behavior (Cowan & Sheridan, 2009). Research has shown the efficacy of adequate teacher behavior management techniques for improving behavioral adjustment and decreasing child EPB, for example by stating clear expectations and rules and consistently using praise (Cowan & Sheridan, 2009). In contrast, inadequate behavior management techniques, such as harsh corrections, have been linked to more child EPB (e.g., Leflot, van Lier, Onghena, & Colpin, 2010).

Based on learning theory, several interventions were developed - although few exclusively focus on teacher behavior management - and found to be effective in reducing child EPB (e.g., Wilson & Lipsey, 2007).

Two-Component Interventions Based on Attachment Theory and Learning Theory

In the parent-child literature, several authors have combined insights from attachment theory and learning theory in two-component interventions, the first component aimed at promoting the quality of the caregiver-child relationship and the second at promoting the caregiver's behavioral management (e.g., Herschell, Calzada, Eyberg, & McNeil, 2002; Van Zeijl et al., 2006). The assumption underlying the combination of these intervention components is that behavioral management, shown to be effective in reducing EPB (e.g.,

Eyberg, Nelson, & Boggs, 2008) will be most effective if it is applied in the context of a warm caregiver-child relationship. Therefore, these interventions start with a child-directed interactive play time. This play time is viewed as a foundation upon which to build a high quality relationship that may promote improved caregiver-child communication, increased positive emotional experiences, and motivation to change within the child (e.g., Barkley, 1987; Driscoll & Pianta, 2010). This, in turn, is expected to facilitate the child's openness for, compliance with, and the efficacy of the behavior management techniques based on learning theory which are practiced during the second intervention component (Barkley, 1987). Parent-Child Interaction Therapy, for example, is a two-component intervention for which the evidence base is large (e.g., Herschell et al., 2002). Van Zeijl and colleagues (2010) found positive effects on child behavior of another attachment and learning theory based intervention promoting "sensitive discipline".

Similarly, school-based interventions combining both components are on the rise. For example, Teacher-Child Interaction Therapy (McIntosh, Rizza, & Bliss, 2000), based on Parent-Child Interaction Therapy, aims at promoting a positive teacher-child relationship and at improving the teacher's behavior management in a first and second 6-week component respectively. In the first component, the child can take the lead during the weekly one-on-one play sessions with the teacher. In the second component, the teacher takes the lead and practices behavior management skills. Teacher-Child Interaction Therapy as a whole and a similar intervention applying the skills to the classroom (i.e., Teacher-Child Interaction Training; Lyon et al., 2009) were found to increase positive teacher-child interactions and child compliance, and reduce child disruptive behavior (Filcheck, McNeil, Greco, & Bernard, 2004; Gershenson et al., 2010; McIntosh et al., 2000; Tiano & McNeil, 2006). Yet, these studies suffer from methodological limitations, such as a lack of a randomized control group

and a small number of participating classes or children. Moreover, to our knowledge, these studies did not explicitly investigate effects on teacher-child relationship quality.

This Study

We investigated the efficacy of an intervention that was built on Teacher-Child Interaction Therapy (e.g., McIntosh et al., 2000) and on Banking Time (e.g., Driscoll & Pianta, 2010), the latter being an elaboration of the first, relationship-focused component of Teacher-Child Interaction Therapy. The intervention, called Playing-2-gether, retained the foundations of both interventions, but adapted them to the Flemish school context, and complemented them with behavior management techniques as described by Cowan and Sheridan (2009). Similar to Teacher-Child Interaction Therapy, Playing-2-gether consists of two 6-week components. Both components stress the importance of teacher-child interactions for preschooler behavioral adjustment. They each focus on different aspects of these teacher-child interactions that are judged to be complementary in preventing EPB. More

specifically, the first component, based on attachment theory, focuses on child-centered play time and is aimed at promoting TCR. The increased positive exchanges and experiences resulting from this high-quality relationship, are, in turn, expected to facilitate the child's openness for and compliance with the behavior management techniques, based on learning theory, practiced during the second intervention component.

The main goal of this study is to investigate the effect of the Playing-2-gether intervention on TCR and on teacher-rated child behavioral adjustment in general. As Playing-2-gether - in particular the first component - is focused on improving TCR and referring to previous research (e.g., McIntosh et al., 2000), we hypothesize that Playing-2-gether will lead to an increase in teacher-child closeness and a decrease in conflict. Moreover, as Playing-2-gether - in particular the second component - explicitly provides the teacher with techniques to decrease child EPB and following previous studies (e.g., Lyon et al., 2009), we hypothesize

that the intervention will lead to an increase in behavioral adjustment in general and a decrease in EPB in particular.

In addition, we intend to explore what happens during this two-component intervention by integrating an intermediate assessment after the first, relationship-focused component. Following the Banking Time studies (Driscoll & Pianta, 2010; Driscoll et al., 2011) and referring to attachment theory (e.g., Verschueren & Koomen, 2012), we expect that an increase in closeness, a decrease in conflict, and an increase in behavioral adjustment is already present after the first component. We also aim to explore unique changes in the outcomes during the second component above and beyond changes during the first one. As the second component, unlike the first, is explicitly focused on behavior management techniques which are based on learning theory and which have been shown to be effective in reducing EPB (e.g., Cowan & Sheridan, 2009), we expect that the second component will add to the effects of the first component on behavioral adjustment. Effects of the second component on the TCR were studied exploratively. First, the TCR may further improve because of the enhanced positive behavioral support (e.g., praise, avoiding criticisms) by the teacher, which is assumed to result from Rule-Game. In the parent-child literature, positive effects of learning theory-based interventions on the parent-child relationship quality have been found (e.g., O'Connor, Matias, Futh, Tantam, & Scott, 2013; Webster-Stratton, 1992). Second, alternatively, the TCR quality may decrease because of the shift in approach (i.e., from child-directed to teacher-directed) and the related increase of behavioral control of the teacher. Third, another alternative would be that the TCR that has been built during the first component of the intervention does not further increase but may be sufficiently strong to withhold the shift in approach (i.e., no change in the TCR).

Method

Participants

In the school year 2009-2010, 46 mainstream schools were recruited. Schools were located in urban areas in the Flemish region of Belgium (Statistics Belgium, 2012). In Flanders, preschool education is available for children aged 2.5 to 6, and is almost always state-funded. Children between the ages of 2.5 and 3 are mostly taught in a separate class, which is a sort of pre-preschool class. The first, second, and third preschool group respectively consist of children aged between 3 and 4 years, between 4 and 5 years, and between 5 and 6 years. Preschool teachers in Flanders are qualified by a Bachelor's degree of education. Although education in Flanders is not obligatory until the age of 6, almost all 3- to 4-year-olds (98.9%) are registered in a preschool (Organization for Economic Cooperation and Development, 2012).

First, parental consent for participation in a screening procedure was requested for 3,747 children of the first and second preschool group of these schools and was obtained for 3,613 (96.4%) children (see flowchart, Figure 1). If parental consent was given, teachers completed a checklist (Behar, 1977) assessing child EPB (M = 1.40, SD = 0.52).

Next, the male preschooler with the highest score on this checklist was selected from each class (see below). We decided to exclusively target boys, as boys, on average show higher EPB levels than girls (e.g., Rutter, Caspi, & Moffitt, 2003) and tend to be more responsive to early teacher-child interactions than girls (e.g., Hamre & Pianta, 2001). In our study, as expected, boys (M = 1.52, SD = 0.58) compared to girls (M = 1.28, SD = 0.39) received higher EPB-scores (t(3144.99) = -14.84, p < .001).

All schools had three or four classes of the first and second preschool group, resulting in 175 participating children.[1] In each classroom, parental consent to participate in the intervention part of the study was requested for the selected male preschooler. If the boy with the highest score on the screening questionnaire was not able to participate, the boy with the second highest score was selected, and so on. Reasons for not being able to participate were

parental refusal (41 children), being absent from school for a long period (two children), or meeting the exclusion criteria (five children).[2] For the 175 participating children, the mean score on the screening questionnaire was 2.24 (SD = 0.69) which is higher than the mean, but lower than a more extreme score of 1.5 SD above the mean (2.39) in the total sample of boys.

In the subsequent school year, the selected children (N = 175) participated in the intervention study, together with their new teacher. Per class, only one teacher-child dyad participated, to ensure feasibility of intervention implementation and data collection. The children's age ranged from 3 years and 9 months to 5 years and 9 months, with an average of 4 years and 9 months (SD = 7 months). Most children (91.4%) had the Belgian nationality. The other children had the Dutch (0.6%), Spanish (0.6%), Turkish (0.6%), or two non-Belgian nationalities (1.1%). For 5.7% of the children, nationality was unknown. The home language was Dutch (i.e., the official language in Flanders) for 70.9% of the children, Dutch and another language for 9.7%, only another language for 17.1%, and unknown for 2.3% of the children

The majority of the children's parents (84% mothers, 85.7% fathers) had the Belgian nationality. Less than half of the parents completed higher education (45.1% mothers, 38.9% fathers). The other parents finished senior high school (35.4% mothers, 34.9% fathers), junior high school (14.3% mothers, 18.2% fathers), or primary school (2.9% mothers, 4.0% fathers). Educational level was unknown for 2.3% of the mothers and 4.0% of the fathers.

Most classes (81.3%) had one full-time teacher. In classes with two part-time teachers (18.7%), the teacher spending the most time in the classroom participated. Because of illness or pregnancy, 13 teachers were replaced during the course of the study. The fill-in teachers participated further with the selected children. The teachers (98.3% female) were on average 39 years old (SD = 9 years) and had on average 17 years (SD = 9 years) of teaching

experience in general education. Four teachers had supplementary training in special education. Intervention status was not related to teacher demographic variables.

At post-test, 166 (94.9%) children still participated in the study. Drop-out was due to school changes (three children), long-term absence of the child (two children) or teacher withdrawal (four children). Drop-out was not significantly related to any of the other measures of TCR or behavioral adjustment at Wave 1, nor to intervention status or preschool group (second or third).

Measures

Teachers completed the Closeness and Conflict subscales of the *Student-Teacher Relationship Scale* (STRS; Pianta, 2001; authorized Dutch translated version by Koomen, Verschueren, & Pianta, 2007). Closeness refers to the openness and warmth in the TCR (11 items, e.g., "I share an affectionate, warm relationship with this child"). Conflict accounts for the degree of discordant teacher-child interactions (11 items, e.g., "This child and I always seem to be struggling with each other"). STRS-subscales are rated on a 5-point scale ranging from 1 (*Definitely does not apply*) to 5 (*Definitely applies*). Subscale scores are calculated by averaging item scores. Evidence for convergent and predictive validity has been found (e.g., Doumen, Koomen, Buyse, Wouters, & Verschueren, 2012; Koomen, Verschueren, van Schooten, Jak, & Pianta, 2012). In our study, mean internal consistency (Cronbach's α) across the three data waves was .87 for Closeness and .88 for Conflict.

Second, teachers filled out the *Strengths and Difficulties Questionnaire* (SDQ; Goodman, 1997; Dutch translation by van Widenfelt, Goedhart, Treffers, & Goodman, 2003). The SDQ consists of five subscales with five items, referring to emotional symptoms (e.g., "Worries", mean α in this study = .64), conduct problems (e.g., "Fights", mean α = .70), hyperactivity/inattention (e.g., "A busy child", mean α = .80), peer relationship problems (e.g., "Solitary", mean α = .66), and prosocial behavior (e.g., "Helps out", mean

α = .78) respectively. As the mean Cronbach's alpha for emotional symptoms is considered to be undesirable and the mean Cronbach's alpha for peer relationship problems is only minimally acceptable (DeVellis, 2003), we decided to exclude these scales from all analyses. All items were scored on a 3-point Likert-scale (0 = *Not true*, 1 = *A little true*, 2 = *Certainly true*). In previous research, the factor structure of the SDQ has been confirmed, and convergent and predictive validity for the subscales were shown (Goodman, 2001; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006). In the following, we refer to the subscales "conduct problems" and "hyperactivity/inattention" as indicators of child EPB.[3]

Family and teacher background were assessed using a parent and a teacher questionnaire respectively.

The Playing-2-gether Intervention

Playing-2-gether is developed for preschoolers with relatively high levels of EPB and their teachers. The program aims at decreasing child EPB through targeting teacher-child interactions. Playing-2-gether consists of two 6-week components during which the teacher organizes one-on-one play sessions with the target child. Play sessions take place a minimum of two times a week, for approximately 15 minutes per session. During the sessions, the teacher practices skills to improve the TCR (first component) and teacher behavior management (second component). The sessions are aimed to provide teachers with a safe environment to practice, but the teachers are also encouraged to apply these skills to the classroom. For example, during the training sessions (see further), teachers and consultants discuss how and when teachers may practice the skills in the classroom.

In the first component, Relationship-Game, the activity is chosen by the child, and the teacher's behavior is constrained to describing the child's activity and conveying understanding. In short, by observing the child during the game, following his lead, imitating his game, narrating his actions and labeling his feelings, the teacher conveys acceptance and

sensitivity to the child (see introduction; Driscoll & Pianta, 2010). When the teacher feels comfortable with these basic techniques, he or she can take the child's relational needs into account and let them guide his/her actions (e.g., if the teacher feels like the child has difficulty expressing his/her emotions, the teacher can label his/her feelings more frequently to give a voice to the child's emotions, cf. "developing relational themes" in Banking Time; Pianta & Hamre, 2001). All the skills mentioned above are intended to make teacher-child interactions more child-centered instead of teacher-centered, which is assumed to improve TCR (e.g., Driscoll & Pianta, 2010).

The second component, Rule-Game, is led by the teacher. The teacher uses skills to increase appropriate child behavior and reduce EPB, such as giving clear commands, introducing rules for good behavior, introducing a pictogram of a kangaroo that urges the child to behave appropriately, and praising the child following good behavior (Cowan & Sheridan, 2009). If EPB still occurs, the teacher can make use of a time-out, but it is important that the atmosphere remains positive.

Teachers in the intervention condition implemented Playing-2-gether according to a standardized teacher manual (Vancraeyveldt, Van Craeyevelt, Verschueren, & Colpin, 2010). Moreover, they were trained and supervised face-to-face on four occasions by Playing-2-gether consultants, either school psychologists or Master's students in school psychology (final year). Consultants were given a consultant manual, including a DVD with several good/bad practice examples of the Playing-2-gether skills (Vancraeyveldt, Van Craeyevelt, Veyt, Verschueren, & Colpin, 2010) and they received seven two-hour sessions of training and supervision by the consultant-coordinator. Consultants were not involved in data collection to maximize independency of data collection (cf. Flay et al., 2005).

At the start of each intervention component, the consultant provided training to the teacher. These training sessions were designed to (a) introduce the teachers to the Playing-2-gether skills, (b) help solving practical problems, such as finding a teacher for supervision of

the other children during the sessions, finding a place and time for the sessions during the school day, and (c) inspire teachers to apply the Playing-2-gether skills in the sessions and in the classroom. In the fourth week of each of the intervention components, a consultation took place, including a recorded observation of a Playing-2-gether session, followed by a video-feedback session with the teacher. Consultation and performance feedback have been shown to improve the outcomes of school-based interventions beyond those achieved through instructional workshops (e.g., Gorman-Smith, 2003). Teacher participation rates in training and consultation sessions ranged between 88.8% and 98.9%.

Teachers in the intervention group completed a diary in which they planned and evaluated each Playing-2-gether session. Based on this diary, we calculated the number of sessions that took place, and used this as a measurement of dosage (range: 0 to 24 sessions). The mean number of completed Relationship-Game sessions was 9.70 (80.8% of the prescribed 12 sessions, *SD* = 1.89), whereas the mean number of completed Rule-Game sessions was 9.38 (78.2% of the prescribed 12 sessions, *SD* = 2.38). This is considered more than sufficient, as positive results of interventions have often been attained with implementation levels around 60% (Durlak & DuPre, 2008). Mean numbers of completed Relationship-Game and Rule-Game sessions did not differ significantly ($t(83) = 1.14$, $p = .26$).

Design and Procedures

A randomized design was used with the teacher-child dyad as unit of randomization. At the beginning of the school year 2010-2011, 175 teacher-child dyads in 46 schools (i.e., 37 schools with four participating classes and nine schools with three participating classes) were randomly assigned per school and per preschool group to an intervention or a control condition, based on an arbitrary criterion. As a result, 89 and 86 teacher-child dyads were assigned to the intervention and control condition respectively.

Data collection took place at three time points. Following the pre-test measurement (Wave 1, October - December 2010), the first 6-week component (i.e., Relationship-Game) was implemented in the intervention condition; the control children received education as usual. Afterwards, an intermediate assessment took place (Wave 2, December 2010 - March 2011), followed by the second 6-week component (i.e., Rule-Game). Again, the control children received education as usual. Next, there was a post-test measurement (Wave 3, February - May 2011).

At Wave 1, parents completed a questionnaire on family background (response rate 97.7%). A questionnaire concerning teacher background, the SDQ and the STRS were sent to the teachers, and assembled after a few weeks. Teacher response rates at Wave 1, 2, and 3 were 98.9%, 94.3%, and 87.4% respectively. There were no significant pre-test differences between intervention and control group on the TCR and behavioral adjustment measures at Wave 1.

Plan for Analysis

As the data reflect a hierarchical structure with measurements nested in children and children in schools, multilevel modeling was used (Peugh, 2010). Model parameters were estimated using the restricted maximum likelihood (REML) procedure for mixed model analysis in PASW Statistics 18 (SPSS Inc., 2009). REML estimation was used because REML in contrast to full maximum likelihood takes uncertainty into account regarding the fixed effect values when the variance parameters are estimated (Raudenbush & Bryk, 2002). Model fit indices such as Model deviance, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were used to compare different models for each outcome variable (Peugh, 2010).

In a first step, unconditional models (i.e., intercept only models) were estimated for all outcome variables to investigate the amount of variance of the scores at the different levels

(i.e., measurement, child, and school level). Modeling the school level was not necessary, because of small intra class correlations ranging from 0 to .04 and small design effects ranging from 1 to 1.12 (which is below the critical value of 2; Muthén & Satorra, 1995). Therefore, the analyses were continued with two-level models. For all outcome variables, a considerable proportion (54% to 73%) of the total variance was located at the child level.

In a second step, several predictors were added to the unconditional models: a dummy-coded Intervention status variable (0 = control group, 1 = intervention group), a continuous, linear Time variable (0 = first wave, 1 = second wave, 2 = third wave), and a two-way interaction between Time and Intervention status. In these models, the dummy-coded Intervention status refers to the difference between intervention and control group in the outcome variables at Wave 1, whereas Time represents the expected linear increase in the outcome variable scores for the control group over time. The two-way interaction between Time and Intervention status represents the effect of Playing-2-gether on the time trend from Wave 1 to Wave 3.

In a third step, discontinuous multilevel modeling (e.g., Singer & Willett, 2003; Wouters, De Fraine, Colpin, Van Damme, & Verschueren, 2012) was used to examine the change in the outcomes after the first and second intervention component respectively. Indeed, the transition from no intervention to Relationship-Game and the transition from Relationship-Game to Rule-Game may result in abrupt (i.e., discontinuous) changes in teacher-child relationship quality and child behavioral adjustment, which are not captured in continuous parameter models. To model these abrupt changes, the two-way interaction term between Time and Intervention status from the second step (see above) was replaced by two two-way interaction terms representing the effects of the two intervention components. First, to represent the effect of Relationship-Game on the change in scores from Wave 1 to Wave 2, a two-way interaction term between Intervention status and a first Dummy variable

representing the second data wave (0 = first wave, 1 = second wave, and 1 = third wave) was added. For the control group, this interaction term is zero. Second, to model the change in the outcome variables following Rule-Game on top of the effect of Relationship-Game, a two-way interaction between Intervention status and a second Dummy variable that represents the third wave (0 = first wave, 0 = second wave, 1 = third wave) was added. It should, however, be noted that we cannot interpret the regression coefficient of this interaction term as a standalone effect of Rule-Game, as the intervention components were not counterbalanced (i.e., Relationship-Game always preceded Rule-Game).

Next to the abovementioned fixed effects, we controlled for random effects in the second and third step of the analyses. Random effects refer to individual deviations from the mean intervention (component) effect. These random effects were only allowed if they improved model fit based on the Model deviance test (Peugh, 2010).

In addition to relying on significant test values, we also computed effect sizes of the intervention effects. To this end, the difference in expected scores between the two groups, divided by the unexplained standard deviation (i.e., the square root of the sum of the unexplained between-person and within-person variance) was calculated. The interpretation of these effect sizes is similar to the interpretation of Cohen's *d* (Cohen, 1992; Van den Noortgate & Onghena, 2008). More specifically, effect sizes of 0.20 (or approaching 0.20) are considered small, 0.50 is considered medium, and 0.80 is considered large.

Results

Descriptive Statistics

Correlations between the outcome variables are presented in Table 1. Small to medium correlations were found among the STRS subscales, among the SDQ subscales, and among the STRS subscales and the SDQ subscales respectively (cf. Cohen, 1988). A significant negative correlation was found between the conflict and closeness subscales of the STRS. The

externalizing problem behavior (EPB-) subscales of the SDQ (i.e., hyperactivity/inattention and conduct problems) were significantly positively interrelated, and significantly negatively linked with prosocial behavior. In general, closeness was significantly negatively associated with the two EPB-subscales and significantly positively to prosocial behavior. In contrast, conflict was significantly positively related to the EPB-subscales and significantly negatively to prosocial behavior.

Mean scores and standard deviations on the SDQ and STRS subscales are also reported in Table 1. In this sample, the average Conduct problems and Hyperactivity/inattention scores at Wave 1 were 0.56 and 1.16 respectively, which were above the reference group mean but below the 80th percentile in the normative sample (Van Leeuwen et al., 2006); the scores of 5.8% and 20.9% of children would fall in the clinical range (i.e., above the 95th percentile). This suggests that our sample on average has relatively high, but no extreme levels of EPB.

Model Estimations

To estimate the number of missing data, we divided the expected number of cases for the outcome variables on the three measurement occasions (175 x 15 = 2625) by the number of missing cases (176). Overall, 6.70% of the data were missing for at least one of the outcome variables. Missing values were calculated using the Expectation-Maximization method, which assumes a distribution for the partially missing data and bases inferences on the likelihood of that distribution. Participants with and without missing data were compared on the outcome variables using Little's (1988) Missing Completely At Random (MCAR) test. This resulted in a normed χ^2 (χ^2/df = 70.464/54) of 1.30, which indicates a good fit between sample scores with and without imputation (normed $\chi^2 < 2$; Bollen, 1989). In this case, missing values do not flaw the results when using the restricted maximum likelihood

procedure for all primary analyses and missing imputation is not deemed to be necessary (Allison, 2003).

First, we estimated multilevel models with Time, and the two-way interaction between Time and Intervention status. This interaction term refers to the effect of the intervention (see Table 2). A significant linear increase in closeness and prosocial behavior for the control condition was found, but no significant change over time for any of the other variables. Moreover, the intervention children, compared to the control children, significantly decreased in conflict, conduct problems, and hyperactivity/inattention, but they did not significantly change in closeness and prosocial behavior. The effect sizes of the significant intervention effects were small or small to medium. Adding random effects did not significantly improve model fit for any of the outcome variables, indicating that the effect of the intervention does not differ over participants.

Second, we estimated discontinuous multilevel models to examine changes in the STRS and SDQ subscales after the first and second intervention component respectively. The effect of the first intervention component is reflected in the two-way interaction between Intervention status and Wave2, whereas the effect of the second intervention component on top of the effect of the first intervention component is reflected in the two-way interaction between Intervention status and Wave3 (see Table 3). The intervention group, compared to the control group, displayed a small but significantly larger increase in closeness and a small but significantly larger decrease in conflict after Relationship-Game. There was no significant change for the intervention group in comparison to the control group in closeness and conflict after applying Rule-Game on top of Relationship-Game. Concerning the SDQ subscales, the intervention group, in comparison to the control group, displayed a small but significantly larger decrease in conduct problems and in hyperactivity/inattention after Relationship-Game. No significant differences between intervention and control group were found for prosocial

behavior after Relationship-Game. After Rule-Game, there were no significant additional differences between intervention and control group for any of the SDQ subscales on top of differences found after Relationship-Game.

Adding random effects did not significantly improve model fit for any of the outcome variables, indicating that the effect of Relationship-Game does not differ over participants. Also, there were no significant interindividual differences in the change of any of the outcome variables after Rule-Game on top of Relationship-Game.

Discussion

In this study, a three-wave randomized design was used to investigate effects of the two-component Playing-2-gether intervention on the teacher-child relationship (TCR) and teacher-reported child behavioral adjustment for children with relatively high levels of externalizing problem behavior (EPB). In addition, the study aimed to increase insight in what happens during the intervention by including an intermediate assessment after the first, relationship-focused component (i.e., Relationship-Game), and by exploring the changes during the second, behavior management-focused component (i.e., Rule-Game). However, only conclusions concerning the stand-alone effects of Relationship-Game could be drawn, as intervention components were not counterbalanced.

Effects of the Two-Component Playing-2-Gether Intervention

This study provides support for positive effects of Playing-2-gether on some, but not all behavioral adjustment variables. More specifically, intervention children compared to control children showed a significant decrease in conduct problems, and hyperactivity/inattention, but no change in prosocial behavior. Comparison with reference samples suggests that boys in our study had relatively high, but no extreme levels of EPB. These findings suggest that Playing-2-gether achieves its main goal: reducing EPB among children with relatively high levels of EPB. In finding a significant decrease in EPB for the

intervention group in a large-scaled randomized controlled trial, this study supports and extends results of previous small-scaled two-component intervention studies (e.g., Lyon et al., 2009). In addition, it mirrors findings from randomized controlled trials with Parent-Child Interaction Therapy, a parallel intervention focused on parent-child interactions (e.g., Herschell et al., 2002). The non-significant effect on prosocial behavior might be explained by the explicit focus on decreasing EPB in Playing-2-gether, which does not include techniques to improve prosocial behavior.

In addition, our study was the first to investigate effects of a two-component intervention on teacher-child relationship quality. In line with its theoretical foundations (e.g., Cowan & Sheridan, 2009; Pianta et al., 2003), Playing-2-gether was shown to reduce teacher-child conflict. No overall Playing-2-gether effects on closeness were found, which might be due to the change from child-centered to teacher-centered interactions from the first to the second component (see further "Two-component versus one-component interventions").

In general, the significant Playing-2-gether effects were small or small to medium at best. Although newly discovered effect sizes should always interpreted with caution and need replication (Ioannidis, 2008), these small effect sizes are consistent with the effect sizes of other indicated interventions targeting child EPB (e.g., Wilson & Lipsey, 2007). Future research may investigate whether the effects are larger for samples with more extreme or clinical levels of EPB. Nevertheless, small intervention effects could have a proportionately greater impact on the population over time, particularly for young children (Mashburn et al., 2008).

The Stand-Alone Effects of a Relationship-focused Intervention

Regarding the stand-alone effects of Relationship-Game, the study revealed that all Playing-2-gether effects (i.e., on conflict, hyperactivity/inattention, and conduct problems) were already present after this first, relationship-focused intervention component. Moreover,

Relationship-Game on its own produced additional effects on the teacher-child relationship quality.

First, Relationship-Game resulted in a small but significant increase in closeness, supporting previous research with Banking Time, the intervention on which Relationship-Game is based (e.g., Driscoll & Pianta, 2010). This finding shows that enhancing child-centered and sensitive teacher skills (e.g., narrating the child's action, labeling the child's feelings) increases open communication and warmth in the teacher-child relationship, thereby supporting premises of attachment theory (e.g., Driscoll & Pianta, 2010; Pianta et al., 2003). This increase in closeness may be important for children's behavioral and school adjustment, as closeness can have a compensatory effect on the further development of EPB in children with initially elevated levels of EPB (e.g., Silver et al., 2005).

Second, as expected by attachment theory (e.g., Pianta et al., 2003), Relationship-Game on its own resulted in a small but significant decrease in teacher-child conflict. However, this finding was not in line with the findings for Banking Time (Driscoll & Pianta, 2010; Driscoll et al., 2011). These different results for conflict across studies might be due to the selection of participants. Whereas in the present study preschoolers with relatively high levels of EPB were selected, Driscoll and Pianta (2010) and Driscoll and colleagues (2011) respectively selected Head Start children for whom the teacher reported more general adjustment concerns and preschoolers at risk because of socio-economic difficulties. Different populations might show different changes in the TCR during a relationship-focused intervention. Future research should investigate this hypothesis into more detail. Nevertheless, finding this decrease in conflict is important, as conflict in preschool has been shown to relate to academic and behavioral outcomes until eighth grade, especially for children with relatively high levels of EPB in preschool and for boys (e.g., Hamre & Pianta, 2001). Our

study gives the first indication that conflict in the TCR can be influenced as early in life as preschool.

Two-Component versus One-Component Interventions

Contrary to our expectations, no significant effects of Rule-Game above and beyond Relationship-Game were found on any of the outcomes. Remarkably, the increase in closeness after Relationship-Game disappeared after Rule-Game and, as a consequence, did not result in an overall Playing-2-gether effect. To date, our study is the first to give at least some insight in what happens during two-component interventions aimed at enhancing teacher-child relationship quality in a first and at improving behavior management in a second component.

On the one hand, our study provides no evidence that Rule-Game added anything beyond Relationship-Game. Moreover, Rule-Game might have been responsible for the attenuation of the Relationship-Game effect on closeness. This may suggest that a relationship-focused intervention may be enough to produce the desired effects for children with relatively high levels of EPB. To examine this possibility, future research should examine whether Relationship-Game in itself is as effective as the combination of Relationship-Game and Rule-Game.

On the other hand, due to the design (i.e., Relationship-Game always preceding Rule-Game), we should be cautious in drawing such conclusions with regard to the non-significant effects of Rule-Game. It is possible, for example, that the first component, regardless of its content, results in positive effects, because children are not used to spending time alone with the teacher. During the second component, the habituation to this one-on-one time might kick in, which might explain the finding of no additional intervention effects.

In the parent-child literature, one small-scaled study (N = 24) with children exhibiting severe EPB has counterbalanced the intervention components of Parent-Child Interaction

Therapy (Eisenstadt, Eyberg, McNeil, Newcomb, & Funderburk, 1993). At the intermediate assessment, parents who started with the behavior management-focused component reported less child noncompliance and disruptiveness than parents who started with the relationship-focused component. At post-test, the former parents compared to the latter reported less conduct problems. In both conditions, however, children moved from outside to within normal limits on compliance, conduct problems, and internalizing problems. No evaluation of the change in parent-child relationship quality was reported in this study.

Although these findings might suggest that a behavior management-focused component is better to start with than a relationship-focused component, it should be noted that this was only one small-scaled study concerning parent-child interactions with children exhibiting serious EPB for more than six months. Moreover, although parent and teachers show similarities as attachment figures or socialization agents (e.g., Verschueren & Koomen, 2012), the affective quality of the parent-child relationship is generally the result of a longer history of interactions and may be harder to influence with a short-term intervention.

Future research should counterbalance the intervention components to capture their unique contributions and to investigate the added value of two-component to one-component interventions. Moreover, future studies with counterbalanced designs might further investigate which type and/or combination of components is best to improve different outcomes. As such, future studies should not only investigate the effect of intervention (components) on child behavioral adjustment and teacher-child relationship quality as outcome variables, but also on other child variables such as engagement (Roorda, Koomen, Spilt, & Oort, 2011) and teacher variables such as behavior management (Leflot et al., 2010). If these future studies assert that both components are equally effective in all outcome variables for all children, teachers may want to choose the component they feel most comfortable with according to their teaching style and according to characteristics of the child. If Relationship-Game and Rule-Game have

differential effects on different outcomes for different types of children (e.g., children with normal versus clinical EPB), then the teacher may choose the intervention component in function of which outcome he/she judges the most pressing to work on. If Rule-Game on its own has no effects at all, however unlikely based on previous research (e.g., Eisenstadt et al., 1993), this component might be deleted and future research should focus on the effects of Relationship-Game.

Limitations

Despite its strengths, this study has a number of limitations. First, teachers were the only evaluators of the intervention effect. Teachers were not blind to the study condition. In addition, as teachers in the intervention condition invested a lot of time and effort in implementing Playing-2-gether, they might be more inclined to perceive changes in the child's behavior regardless of whether the behavior really changed. This is a difficult challenge in school mental health research because teachers are often in the best role to provide the intervention and assess the participants, especially for young children. To complement our findings, more objective assessments of intervention effects, especially on child EPB, is recommended for future research. On the other hand, finding an effect on teachers' perceptions is valuable on its own, as teachers' perceptions have been shown to predict school-related outcomes such as behavioral engagement and school trajectories, above and beyond objective test results and observations (e.g., Doumen et al., 2012). As several well-known studies point out, teacher perceptions may become self-fulfilling prophecies, although these effects are generally small and teacher expectations predict child outcomes generally more because they are accurate rather than because they are self-fulfilling (Jussim & Harber, 2005). Finally, the use of teacher ratings in itself cannot fully explain the differential effects for different outcomes and intervention components. For example, if teacher rater bias

would be the sole explanation for the intervention effects, it is hard to explain why EPB did not decrease during Rule Game.

A second limitation is the lack of data on implementation quality. In combination with the study design (i.e., Relationship-Game preceding Rule-Game), the lack of these data complicates the evaluation of the relative contribution of both intervention components. Indeed, we cannot exclude that the stronger effects of Relationship-Game are due to teachers' better implementation of these skills compared with Rule-Game skills. The one-on-one sessions may lend themselves better to play than to discipline. Future research should focus on the assessment of both implementation dosage and quality during the one-on-one sessions and in the classroom.

Third, as the internal consistency of the SDQ-subscales peer relationship problems and emotional symptoms was low, these subscales were excluded from the analyses. As a result, the study presents a more narrow view on behavioral adjustment than was intended.

Fourth, generalizability of the findings is limited due to sample selection. The sample consisted of only male preschoolers. Future research should investigate whether these findings also hold for female preschoolers, who tend to exhibit lower EPB levels (e.g., Rutter et al., 2003) and, on average, seem less responsive to the effects of early teacher-child interactions (Hamre & Pianta, 2001). Moreover, future studies should also examine the Playing-2-gether intervention in populations with clinical levels of EPB, and examine if and how intervention outcomes may differ depending on the presence of child psychiatric disorders or other problems (e.g., Attention Deficit Hyperactivity Disorder or Oppositional Defiant Disorder, depression, anxiety, problems at home).

Fifth, follow-up measurements would allow us to investigate whether Playing-2-gether effects on conflict and child EPB are maintained in the long run. Moreover, follow-up research would shed light on possible sleeper effects (i.e., delayed effects; Barnett, 2011).

Indeed, TCR quality and EPB in preschool have been shown to be predictive of academic and behavioral outcomes later on, especially for children showing EPB (e.g., Dodge et al., 2006; Hamre & Pianta, 2001).

Conclusions

This study showed that the Playing-2-gether intervention reduced teacher-child conflict and child externalizing behavior, and thereby contributes to the empirical evidence on two-component interventions targeting teacher-child interactions to decrease child externalizing problem behavior. Additionally, this study is the first to shed some light on the (differential) contribution of a relationship-focused and a behavior management-focused intervention component in the effects of two-component teacher-child interventions. Our findings showed that a first, relationship-focused intervention component, in itself, produced positive effects on child externalizing problem behavior, closeness, and conflict.

To further disentangle the unique effects of both components, future research should reverse the order of the intervention components and/or compare the effects of two-component-interventions with one-component-interventions. Moreover, future studies are strongly recommended to include intermediate assessments to gain more insight in what happens during the intervention.

References

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112,* 545-557. doi: 10.1037/0021-843X.112.4.545

Barkley, R. (1987). *Defiant children: A clinician's manual for parent training.* New York: Guilford Press.

Barnett, W. S. (2011). Effectiveness of early childhood education. *Science, 333,* 975-978. doi: 10.1126/science.1204534

Behar, L. B. (1977). Preschool Behavior Questionnaire. *Journal of Abnormal Child Psychology, 5,* 265-275. doi: 10.1007/BF00913697

Bollen, K. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, B. (2008). Classroom problem behavior and teacher-child relationships in kindergarten: The moderating role of classroom climate. *Journal of School Psychology, 46,* 367-391. doi: 10.1016/j.jsp.2007.06.009

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi: 10.1037/0033-2909.112.1.155

Cowan, R. J., & Sheridan, S. M. (2009). Evidence-based approaches to working with children with disruptive behavior. In T. B. Gutkin & C. R. Reynolds (Eds.), *The Handbook of School Psychology* (pp. 569-590). New York: John Wiley & Sons.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon, R. M. Lerner (Series Eds.), & N. Eisenberg (Vol. Ed.), *Handbook of*

child psychology: Vol. 3, Social, emotional, and personality development (6th ed., pp. 719-788). New York: John Wiley & Sons.

Doumen, S., Koomen, H., Buyse, E., Wouters, S., & Verschueren, K. (2012). Teacher and observer views on student-teacher relationships: Convergence across kindergarten and relations with student engagement. *Journal of School Psychology, 50,* 61-76. doi: 10.1016/j.jsp.2011.08.004

Doumen, S., Verschueren, K., Buyse, E., Germeijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher-child conflict and aggressive behaviour in kindergarten: A three-wave longitudinal study. *Journal of Clinical Child and Adolescent Psychology, 37*, 588-599. doi: 10.1080/15374410802148079

Driscoll, K. C., & Pianta, R. C. (2010). Banking Time in Head Start: Early efficacy of an intervention designed to promote supportive teacher-child relationships. *Early Education and Development, 21,* 38-64. doi: 10.1080/10409280802657449

Driscoll, K. C., Wang, L., Mashburn, A. J., & Pianta, R. C. (2011). Fostering supportive teacher-child relationships: Intervention implementation in a state-funded preschool program. *Early Education and Development, 22,* 593-619. doi: 10.1080/10409289.2010.502015

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327-350. doi: 10.1007/s10464-008-9165-0

Eisenstadt, T. H., Eyberg, S., McNeil, C. B., Newcomb, K., & Funderburk, B. (1993). Parent-Child Interaction Therapy with behavior problem children – Relative effectiveness of two stages and overall treatment outcome. *Journal of Clinical Child Psychology, 22,* 42-51. doi: 10.1207/s15374424jccp2201_4

Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Psychology, 37,* 215-237. doi: 10.1080/15374410701820117

Fearon, R. P., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., Lapsley, A. M., & Roisman, G. I. (2010). The significance of insecure attachment and disorganization in the development of children's externalizing problem behavior: A meta-analytic study. *Child Development, 81*, 435-456. doi: 10.1111/j.1467-8624.2009.01405.x

Filcheck, H. A., McNeil, C. B., Greco, L. A., & Bernard, R. S. (2004). Using a whole class token economy and coaching of teacher skills in a preschool classroom to manage disruptive behavior. *Psychology in the Schools, 41,* 351-361. doi: 10.1002/pits.10168

Flay, B. R., Biglan, A., Boruch, R., González Castro, F., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151-175. doi: 10.1007/s11121-005-5553-y

Gershenson, R. A., Lyon, A. R., & Budd, K. S. (2010). Promoting positive interactions in the classroom: Adapting Parent-Child Interaction Therapy as a universal prevention program. *Education and Treatment of Children, 33,* 261-287. doi: 10.1177/0145445509344215

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586. doi: 10.1111/j.14697610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1337-1345. doi: 10.1097/00004583-200111000-00015

Gorman-Smith, D. (2003). Effects of teacher training and consultation on teacher behavior toward students at high risk for aggression. *Behavior Therapy, 34*, 437-452. doi: 10.1016/S0005-7894(03)80029-1

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625-638. doi: 10.1111/1467-8624.00301

Herschell, A. D., Calzada, E. J., Eyberg, S. M., & McNeil, C. B. (2002). Parent-Child Interaction Therapy: New directions in research. *Cognitive and Behavioral Practice, 9*, 9-15. doi: 1077-7229/02/9-1651.00/0

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19,* 640-648.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9,* 131-155. doi: 10.1207/s15327957pspr0902_3

Koomen, H., Verschueren, K., & Pianta, R. (2007). *Leerling-Leerkracht Relatie Vragenlijst (LLRV): Handleiding. [Student-Teacher Relationship Scale: Manual]*. Houten, The Netherlands: Bohn Stafleu van Loghum.

Koomen, H. M. Y., Verschueren, K., van Schoten, E., Jak, S., & Pianta, R. C. (2012). Validating the Student-Teacher Relationship Scale: Testing factor structure and measurement invariance across child gender and age in a Dutch normative sample. *Journal of School Psychology, 50,* 215-234. doi: 10.1016/j.jsp.2011.09.001

Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behavior: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology, 38,* 869-882. doi: 10.1007/s10802-010-9411-4

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198–1202. doi: 10.2307/2290157

Lyon, A. R., Gershenson, R. A., Farahmand, F. K., Thaxter, P. J., Behling, S., & Budd, K. S. (2009). Effectiveness of Teacher-Child Interaction Training (TCIT) in a preschool setting. *Behavior Modification, 33,* 855-884. doi: 10.1177/0145445509344215

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79,* 732-749. Doi: 10.1111/j.1467-8624.2008.01154.x

McIntosh, D. E., Rizza, M. G., & Bliss, L. (2000). Implementing empirically supported interventions: Teacher-Child Interaction Therapy. *Psychology in the Schools, 37,* 453-462. doi: 10.1002/1520-6807(200009)37:5<453::AID-PITS5>3.0.CO;2-2

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological Methodology* (pp. 267-316). Washington, DC: American Sociological Association.

O'Connor, T. G., Matias, C., Futh, A., Tantam, G., & Scott, S. (2013). Social learning theory parenting intervention promotes attachment-based caregiving in young children: Randomized clinical trial. *Journal of Clinical Child and Adolescent Psychology, 42,* 358-370. doi: 10.1080/15374416.2012.723262

Organisation for Economic Cooperation and Development (OECD) (2012). *Education at a Glance: OECD indicators. Indicator C1: Who participates in education?* Retrieved from http://www.oecd-ilibrary.org/docserver/download/9612031ec023.pdf?expires=1369729041&id=id&accname=ocid49025850&checksum=DE24D0267F05EB7DC5A7B405B394B54A

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 46,* 85-112. doi: 10.1016/j.jsp.2009.09.002

Pianta, R. C. (2001). *Student–Teacher Relationship Scale: Professional manual.* Lutz, FL: Psychological Assessment Resources.

Pianta, R. C., & Hamre, B. (2001). *Students, teachers, and relationship support [STARS]: User's guide.* Lutz, FL: Psychological Assessment Resources.

Pianta, R. C., Hamre, B. K., & Stuhlman, M. (2003). Relationships between teachers and children. In W. M. Reynolds, G. E. Miller, & I. B. Weiner (Eds.), *Handbook of Psychology. Vol. 7: Educational psychology* (pp. 199-234). New York: John Wiley & Sons.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, Second edition.* Newbury Park, CA: Sage

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81,* 493-529. doi: 10.3102/0034654311421793

Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *Journal of Child Psychology and Psychiatry, 44,* 1092-1115. doi: 10.1111/1469-7610.00194

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing problem behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology, 43,* 39-60. doi: 10.1016/j.jsp.2004.11.003

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Smidts, D. P., & Oosterlaan, J. (2007). How common are symptoms of ADHD in typically developing preschoolers? A study on prevalence rates and prenatal/demographic risk factors. *Cortex, 43,* 710-717. doi: 10.1016/S0010-9452(08)70500-8

SPSS Inc. (2009). *PASW Statistics 18.0 Command Syntax Reference.* Chicago: SPSS Inc.

Statistics Belgium (2012). *Typologie van gemeenten volgens twee verschillende bronnen: Eurostat en OESO* [Typology of areas according to two sources: Eurostat and OECD]. Brussels, Belgium: Statistics Belgium.

Tiano, J. D., & McNeil, C. B. (2006). Training Head Start teachers in behavior management using Parent-Child Interaction Therapy: A preliminary investigation. *Journal of Early and Intensive Behavioral Intervention, 3,* 220-233. Retrieved from http://www.baojournal.com/JEIBI/JEIBI-VOL-3/JEIBI-3-2.pdf

Vancraeyveldt, C., Van Craeyevelt, S., Verschueren, K., & Colpin, H. (2010). *Samen-Spel in de kleuterklas. Leerkrachthandleiding* [Playing-2-gether in preschool. Teacher manual]. Leuven: KU Leuven, School Psychology and Child and Adolescent Development.

Vancraeyveldt, C., Veyt, T., Van Craeyevelt, S., Verschueren, K., & Colpin, H. (2010). *Samen-Spel in de kleuterklas. Begeleiderhandleiding* [Playing-2-gether in preschool. Consultant manual]. Leuven: KU Leuven, School Psychology and Child and Adolescent Development.

Van Den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2,* 142-151.

Van Leeuwen, K., Meerschaert, T., Bosmans, G., De Medts, L., & Braet, C. (2006). The Strengths and Difficulties Questionnaire in a community sample of young children in Flanders. *European Journal of Psychological Assessment, 22,* 189-197. doi: 10.1027/1015-5759.22.3.189

van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry, 12,* 281-289. doi: 10.1007/s00787-003-0341-3

Van Zeijl, J., Mesman, J., van Ijzendoorn, M. H., Bakermans-Kranenburg, M. J., Juffer, F., Stolk, M. N., & Koot, H. M. (2006). Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing problem behavior problems: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 994-1005. doi: 10.1037/0022-006X.74.6.994

Verschueren, K., & Koomen, H. M. Y. (2012). Teacher-child relationships from an attachment perspective. *Attachment and Human Development, 14,* 205-211. doi: 10.1080/14616734.2012.672260

Webster-Stratton, C. (1992). Individually administered videotape parent training: "Who benefits?". *Cognitive Therapy and Research, 16,* 31-35.

Webster-Stratton, C., & Reid, M. J. (2009). Parents, teachers and therapists using the child-directed play therapy and coaching skills to promote children's social and emotional competence and to build positive relationships. In C. E. Schaefer (Ed.), *Play therapy for preschool children* (pp. 245-273). Washington, DC: American Psychological Association.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. Update of a meta-analysis. *American Journal of Preventive Medicine, 33,* S130-S143. doi: 10.1016/j.amepre.2007.04.011

Wouters, S., De Fraine, B., Colpin, H., Van Damme, J., & Verschueren, K. (2012). The effect of track changes on the development of academic self-concept in high school: A dynamic test of the Big-Fish-Little-Pond effect. *Journal of Educational Psychology, 104,* 793-805. doi: 10.1037/a0027732

Footnotes

[1] The total number of participating children (N = 175) was chosen based on simple prior power calculations, showing that when using a between-subject design when comparing two mean score levels at a .05 significance level, 175 students are sufficient to have a high probability to detect a small to moderate effect (power > .80 for d = 0.40).

[2] The exclusion criteria for child participation in the intervention were: (a) a diagnosis of autism spectrum disorder, (b) a suspicion of autism spectrum disorder as judged by the child's teacher and/or parent(s), and/or (c) being enrolled at a similar intervention for behavioral or emotional problems at the time of the screening.

[3] The subscale hyperactivity/inattention may be considered as a hybrid scale and not a pure measure of EPB, as it consists of items referring to both hyperactivity and attention. Therefore, we deleted the item of the SDQ which refers the most to inattention ("Easily distracted, concentration wanders"), and we conducted the main analyses in the paper with the four remaining items of this "pure" measure of hyperactivity. The results of the analyses with and without this inattention item were similar. Therefore, in this manuscript, we report on the results with the original hyperactivity/inattention scale. The output of these analyses is available on request.

Figure 1

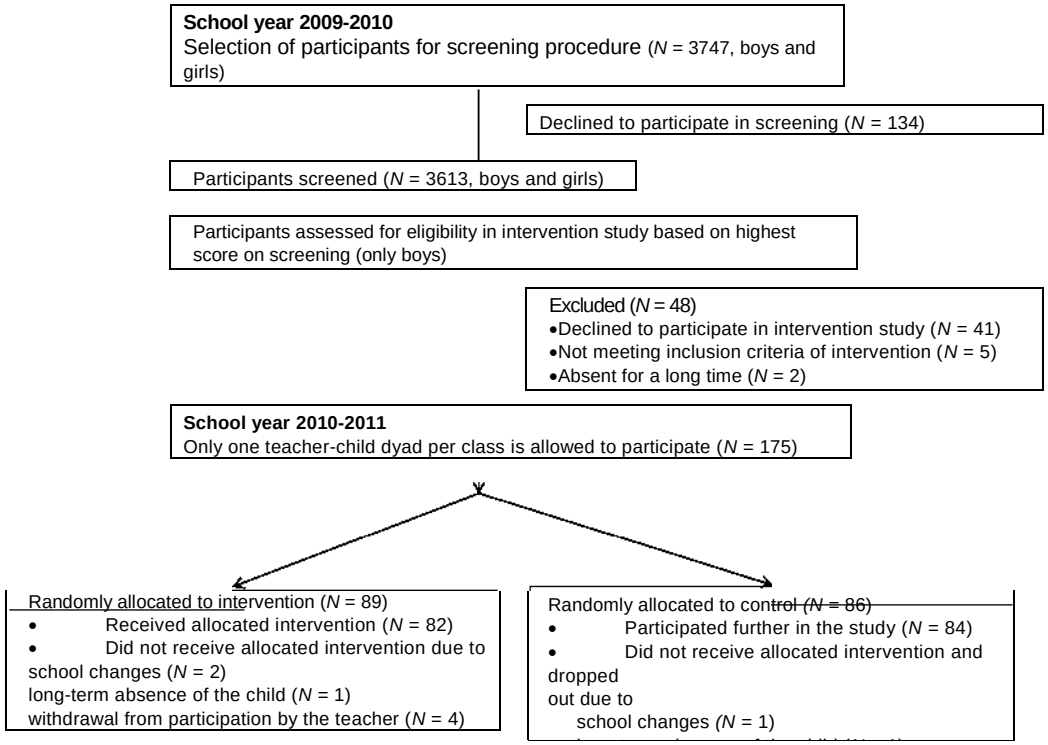

Figure1 Flowchart of classes and participants in the randomized control trial

Tablet

Table 1

Bivariate Correlations, Means and Standard Deviations of the STRS and the SDQ Subscales for the Total Sample (N = 175), and Means and Standard Deviations for the Intervention (n = 89) and Control Group (n = 86) Separately

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.CLOSE1	1														
2.CLOSE2	.66**	1													
3.CLOSE3	.63**	.72**	1												
4.CONFLICT1	-.27**	-.25**	-.16*	1											
5.CONFLICT2	-.21**	-.40**	-.31**	.72**	1										
6.CONFLICT3	-.16*	-.29**	-.42**	.68**	.81**	1									
7.COND1	-.21**	-.28**	-.25**	.59**	.59**	.54**	1								
8.COND2	-.15	-.40**	-.35**	.53**	.53**	.64**	.66**	1							
9.COND3	-.11	-.27**	-.38**	.40**	.40**	.62**	.64**	.76**	1						
10.PROS1	.57**	.41**	.50**	-.31**	-.31**	-.27**	-.39**	-.34**	-.25**	1					
11.PROS2	.42**	.50**	.55**	-.29**	-.29**	-.40**	-.35**	-.50**	-.41**	.69**	1				
12.PROS3	.41**	.44**	.63**	-.23**	-.23**	-.48**	-.29**	-.42**	-.48**	.54**	.71**	1			
13.HYP/INAT1	-.21**	-.12	-.10	.31**	.31**	.33**	.35**	.25**	.22**	-.36**	-.37**	-.27**	1		
14.HYP/INAT2	-.04	-.17*	-.13	.38**	.38**	.47**	.37**	.43**	.31**	-.24**	-.38**	-.30**	.68**	1	
15.HYP/INAT3	-.07	-.18*	-.17*	.24**	.24**	.35**	.31**	.32**	.33**	-.20*	-.35**	-.33**	.70**	.78**	1
M total sample	3.69	3.86	3.98	1.91	1.87	1.80	0.56	0.51	0.42	1.03	1.09	1.15	1.16	1.05	1.01
SD total sample	0.66	0.67	0.67	0.74	0.82	0.71	0.43	0.44	0.41	0.49	0.52	0.51	0.58	0.58	0.55
M control	3.72	3.81	3.95	1.90	1.98	1.90	0.61	0.62	0.52	1.04	1.04	1.13	1.15	1.10	1.10
SD control	0.64	0.67	0.68	0.74	0.82	0.71	0.47	0.47	0.45	0.50	0.51	0.49	0.59	0.60	0.57
M intervention	3.66	3.91	4.01	1.91	1.77	1.70	0.51	0.40	0.33	1.02	1.14	1.17	1.16	1.00	0.91
SD intervention	0.69	0.67	0.66	0.73	0.81	0.71	0.38	0.38	0.35	0.48	0.53	0.53	0.58	0.57	0.51

Note. CLOSE = Closeness at Waves 1, 2, and 3 (CLOSE1, CLOSE2, CLOSE3), COND = Conduct problems. PROS = Prosocial Behavior. HYP/INAT = Hyperactivity/Inattention.
*$p < .05$. **$p < .01$.

Table 2

Parameter Estimates and Corresponding Standard Errors for the Multilevel Models for the Overall Effect of the Playing-2-Gether Intervention Over Three Waves (N = 175)

Parameter	CLOSENESS	CONFLICT	CONDUCT	HYP/INAT	PROSOCIAL
Fixed effects					
Intercept (initial status)	3.69 (.05)***	1.91 (.06)***	0.57 (.03)***	1.15 (.04)***	1.02 (.04)***
Time	0.12 (.03)***	0.01 (.03)	-0.03 (.02)	-0.03 (.02)	0.05 (.02)*
Intervention status * Time	0.07 (.04)	-0.11 (.04)**	-0.06 (.02)*	-0.10 (.03)**	0.02 (.03)
Effect size *(d)*	0.21	-0.29	-0.28	-0.35	0.08
Random effects					
Intercept (Level 2)	0.30 (.04)***	0.41 (.05)***	0.13 (.02)***	0.23 (.03)***	0.17 (.02)***
Residual (Level 1)	0.15 (.01)***	0.15 (.01)***	0.06 (.00)***	0.09 (.01)***	0.09 (.01)***
Deviance	797.41	856.17	341.01	594.79	533.93
AIC	801.41	860.17	345.01	598.79	537.93
BIC	809.79	868.55	353.39	607.15	546.31
Parameters	5	5	5	5	5

Note. CONDUCT = Conduct problems. HYP/INAT = Hyperactivity/Inattention. PROSOCIAL= Prosocial behavior. Standard errors are in parentheses. Effect size *(d)* = An adaptation of Cohen's *d* (Cohen, 1992; Van den Noortgate & Onghena, 2008).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3

Parameter Estimates and Corresponding Standard Errors for the Discontinuous Multilevel Models for the Effect of the Two Components of the Playing-2-Gether Intervention over Three Waves (N = 175)

Parameter	CLOSENESS	CONFLICT	CONDUCT	HYP/INAT	PROSOCIAL
Fixed effects					
Intercept (initial status)	3.68 (.05)***	1.92 (.06)***	0.57 (.03)***	1.16 (.04)***	1.01 (.04)***
Time (linear)	0.12 (.03)***	0.01 (.03)	-0.03 (.02)	-0.03 (.02)	0.05 (.02)*
Intervention status * Wave2	0.12 (.06)*	-0.13 (.06)*	-0.09 (.04)*	-0.13 (.05)**	0.06 (.05)
Effect size *(d)*	0.18	-0.17	-0.21	-0.23	0.12
Intervention status * Wave3	-0.00 (.07)	-0.08 (.07)	-0.03 (.04)	-0.05 (.05)	-0.03 (.05)
Effect size *(d)*	-0.00	-0.11	-0.07	-0.09	-0.06
Random effects					
Intercept (Level 2)	0.30 (.04)***	0.41 (.05)***	0.12 (.02)***	0.24 (.03)***	0.17 (.02)***
Residual (Level 1)	0.15 (.01)***	0.15 (.01)***	0.06 (.00)***	0.09 (.01)***	0.09 (.01)***
Deviance	798.60	858.60	343.55	597.10	535.89
AIC	802.60	862.60	347.55	601.10	539.89
BIC	810.97	870.97	355.93	609.48	548.25
Parameters	6	6	6	6	6

Note. CONDUCT = Conduct problems. HYP/INAT = Hyperactivity/Inattention. PROSOCIAL= Prosocial behavior. Standard errors are in parentheses. Effect size *(d)* = An adaptation of Cohen's *d* (Cohen, 1992; Van den Noortgate & Onghena, 2008).
*$p < .05$. **$p < .01$. ***$p < .001$.